ELECTRONIC RECORDS ARCHIVES

SUMMARY REPORT – DIALOGUE SESSIONS (SRDS)

(TOMP v2.0, TASK 4.3.8)

for the

**NATIONAL ARCHIVES AND
RECORDS ADMINISTRATION**

**ELECTRONIC RECORDS ARCHIVES
PROGRAM MANAGEMENT OFFICE
(NARA ERA PMO)**

**Final
April 11, 2003**

Prepared by:

*Integrated Computer Engineering (ICE)
A Subsidiary of
American Systems Corporation (ASC)*

**Contract Number: GS-35F-0673K
Delivery Order Number: NAMA-01-F-0031/02-107**

Final

Table Of Contents

List of Tables

♦ **National Archives and Records Administration** ♦

SUMMARY REPORT – DIALOGUE SESSIONS

1.0   EVENT

The National Archives and Records Administration (NARA) conducted five User Dialogue Sessions in February and March 2003 to meet with ERA system stakeholders from the private and public sector.

1.1   EVENT OBJECTIVES

The objectives of the dialogue sessions were to provide an overview and current update on NARA progress with ERA to the various regional areas and to identify potential requirements for the ERA system through facilitated group discussions with the dialogue session attendees.

1.2   EVENT STRUCTURE

The sessions were designed to allow the attendees to offer their insight to the potential needs and expectations of ERA users. The input from the sessions was then analyzed to identify potential requirements for the ERA system that may have been overlooked.

Each session was structured into three main sections:

1. Overview Presentation on ERA
    -   Overview of the overall challenges and timelines facing NARA and introduction to the ERA system offerings for different user groups – originators and researchers.

2. Breakout Session
    -   The attendees could select to participate in one of two facilitated discussions during the breakout sessions. The producer group discussed the issues facing record originators and the researcher group discussed the issues around accessing the information from ERA.

3. Final Summation
    -   The facilitators covered the main points discussed during each facilitated session and opened the floor to any final comments or suggestions from the attendees.

The Washington DC session was held during a Bi-monthly Records Information Discussion Group (BRIDG) meeting and included only the overview presentation and a short question and answer session.

1.3   EVENT LOCATION

After evaluating the database of the potential users invited to the ERA User Conference held in October of 2002 at the Archives II facility, appropriate locations were selected for the dialogue sessions. **Table 1-1, Event Locations and Times**, provides a list of those locations.

Final

Location	Date	Time
Atlanta, GA (Carter Library)	February 11, 2003	8:30 AM - 12:30 PM
New York, NY (Northeast Region)	February 13, 2003	8:30 AM - 12:30 PM
Laguna Miguel, CA (Reagan Library)	February 25, 2003	8:30 AM - 12:30 PM
Chicago, IL (Great Lakes Region)	February 27, 2003	8:30 AM - 12:30 PM
Washington DC (BRIDG Meeting)	March 20, 2003	1:00 PM - 2:00 PM

Table 1-1: Event Locations and Times

1.4   EVENT MESSAGES

At each event the facilitators delivered the following messages:

- NARA is here to listen to you and to discuss requirements for ERA;
- NARA wants your input; and
- Not building ERA is not an option.

1.5   AUDIENCE

The ERA PMO mailed over 1,300 invitations nationwide to the groups and persons invited to attend the Dialogue Sessions. In addition, the announcement was posted on NARA listservs and emailed to key stakeholders at the NARA Presidential and Regional facilities.

A total of 141 individuals attended across the five dialogue sessions, with large turnouts in both New York (36 attendees) and Chicago (40 attendees). Of the total dialogue session's attendees who submitted evaluation forms 50% indicated they were Federal Agency personnel and 19% indicated they were government employees. Other groups represented included international civil servants, private archivists, educators, genealogists, and Information Managers.

The main sources that notified attendees of the dialogue sessions in their region, were the listservs (30%), emails from colleagues (30%) and the mailing/NARA web site (20%). The Chicago session was well represented due to the efforts of the NARA regional officer directly contacting his major local clients. A list of attendees broken down by dialogue session (except for Washington DC) is available in **Appendix A, Summary of Dialogue Session Attendee Categories.**

1.6   NARA REPRESENTATIVES

Table 1-2, NARA Representatives, details the presenters and facilitators for each dialogue session.

⧫ **National Archives and Records Administration** ⧫

Location	Presenter(s)	Producer's Facilitators	Researcher's Facilitators	Date
Atlanta, Georgia	Adrienne Woods Fynnette Eaton	Dyung Le Fynnette Eaton Michael Skipper	Greg Bluher Adrienne Woods Lynn Bernard	2/11/03
New York, New York	Fynnette Eaton Adrienne Woods	Dyung Le Fynnette Eaton James McAlpin	Greg Bluher Adrienne Woods Carmen Colon	2/13/03
Los Angeles, California	Fynnette Eaton	Dyung Le Fynnette Eaton Carmen Colon	Greg Bluher Richard Steinbacher Lynn Bernard	2/25/03
Chicago, Illinois	Adrienne Woods	Dyung Le Lawrence Bewer James McAlpin	Greg Bluher Adrienne Woods Carmen Colon	2/27/03
Washington DC	Fynnette Eaton Adrienne Woods	Dyung Le Greg Bluher Carmen Colon	Dyung Le Greg Bluher Carmen Colon	3/20/03

Table 1-2: NARA Representatives

2.0   OBSERVATIONS

The Dialogue Sessions proved to be valuable in expanding our understanding of user concerns and priorities. The majority of attendees stated that the Dialogue Session met their expectations and many positive comments were made with respect to fact that NARA was making the effort to keep the users informed and listening to their input about ERA and its long term objectives.

2.1   DIALOGUE SESSIONS FACILITATED BREAKOUT SESSIONS

The Dialogue Session Facilitator notes from each of the breakout sessions are provided in **Appendix B: Dialogue Session Notes**. A summary of the major points raised during the breakout sessions is provided in **Section 3.0**.

2.2   EVALUATION RESULTS

The Evaluation Forms collected from each dialogue session showed a very positive response from the attendees regarding the accomplishment of the session goals and objectives, and the knowledge and communications skills of the presenters and facilitators. Approximately 90% of the attendees stated that the sessions met their expectations, 95% were satisfied with the facilities, and 100% indicated the staff was helpful and pleasant.

⬥ **National Archives and Records Administration** ⬥

2.3   FOLLOW-UP COMMUNICATIONS

As a follow-up to the Dialogue Sessions, an email will be sent to thank the attendees for their participation and inform them that a summary report of the Dialogue Sessions will be posted to the ERA web site, following the publication of the Draft Request for Proposal (DRFP).

2.4   LESSONS LEARNED

The ERA PMO held a Lessons Learned meeting after the first two Dialogue sessions to gather feedback and learn where improvements can be made for future conferences. The lessons learned from the initial meeting, which were incorporated into the subsequent sessions, were:

- Grant approval to publish the attendees information; and
- Ask attendees to hold questions until after the general presentation as all questions would be captured and answered during the breakout sessions.

2.5   NEXT STEPS

The Dialogue Session Summary Report will be reviewed by the Requirements Team for possible inclusion in the ERA Requirements Document. There is some discussion on continuing the Dialogue Sessions to evaluate the ERA Design during the development phase.

3.0   SUMMARY OF FACILITATED DISCUSSIONS

The attendees were allowed to select which breakout session to participate in, selecting between a discussion on the front end of the ERA system (producers) and the back end of the ERA system (researchers). The main system requirements raised by each group across the four sessions are detailed below. As a considerable amount of the discussion touched on areas outside of requirements, a third section is included that covers the attendees' main concerns and issues they foresee with the development of ERA.

Attendees raised the following main points at all four sessions:

- There are serious outstanding issues in respect to electronic records management and archiving;
- A very limited number of agencies have any type of system in place for electronic records management or archiving;
- There are still individuals that deal entirely in paper records and have yet to transition to any type of electronic records; and
- Efficient electronic records management and archiving will require both clear directives and standards for agencies and individuals to follow.

3.1   PRODUCER SUMMARY

The major finding from the facilitated sessions with producers was the belief that for ERA to succeed, NARA had to be the authority on setting the standards for managing and archiving

electronic records. Several attendees believed that standardization is the key to the long-term success of ERA. The major requirement raised by the attendees was for ERA to maintain the functionality of the archived data, as well as the content and context.

There was a clear indication that many agencies and individuals are not clear on what predetermines an electronic record as permanent and that NARA would need to take the lead to provide extensive training in the management of electronic records before any ERA system could be used efficiently.

A summary of notes taken from the producer facilitated discussions is given below. The complete notes from all sessions can be reviewed in **Appendix B, Dialogue Session Notes.**

- NARA needs to set standards (be the authority) – standardization is key
    - Digital photography – process of captioning and metadata and standards for format
    - Clarification of requirements for transfer to NARA
    - Major concern is that agencies will develop their own systems and standards
    - With DOD standards, software vendors have jumped onto the standards – NARA should try to exploit this concept in setting standards
    - Many of the solutions will not work unless users are forced to use them
    - Federal Agencies need to comply to the standards
    - Cost justification – should be cheaper for NARA to go out and develop an all encompassing system than for each agency to develop its own system
    - A centralized system requires an agency-unified approach

- Metadata
    - Organized by keywords, file number, subject
    - Biggest problem is not to define metadata but to get users to enter them
    - NARA needs to inform the user of the benefit of having metadata items organized and the value of having this information at their fingertips
    - Need clear incentive for record originators to use metadata
    - Records must be hierarchically structured and searchable by metadata. We want a logical map of your holdings

- Record Management/Information Technology Relationship
    - Key is to get IT management to embrace record management
    - Need to work with Record Officers and IT officers and ensure they are talking the same language to each other
    - Individuals have no concept of the archiving "schedule"
    - At the office level, most have no idea of how to manage and archive electronic records using file directories, hierarchies etc
    - Everybody is a records manager as we all have access to the "delete" button

- Preservation
    - A record is not useful unless we can access it without losing the functionality of the record, the metadata, and the content

- Need both the code for the operating system and the bit-stream data for Excel – the Excel data is meaningless without access to all the applications tools
- System must retain utility of electronic records (e.g., Excel macro and formula)
- The records that are permanent need to be predetermined. Impact on workflow policy. User education is essential as they do not know what to keep and what is junk
- When we view a document many years from now, what are we going to compare its current presentation with to ensure authenticity?
- The data supporting the initial authentication of a record must be preserved as part of the record to enable future litigation support

- Email and Attachments
  - The content of the email should dictate if it is a record to be kept or to be deleted
  - The attachment must be preserved – both formula as well as format.
  - What is relevant within the attachment – the image, the formulas, ability to change it?
  - Cannot keep all the functionality of the attachment — what is acceptable?
  - Both evidential value and informational value need to be preserved
  - Need to preserve the chain of emails

- Training
  - Individuals/agencies need training to determine which electronic records are permanent and thus need to be archived
  - NARA needs to provide training to the producers in using the system

- The technology for transferring records to NARA needs to work from bottom up, not top down.

- Requirements to scale down as well as up to address more limited need of agencies and State archives

- Migration functionality
  - No value (in term of authenticity) in preserving the old form.
  - Data can be re-manipulated by business rules that must be preserved

- Administrative use of databases – more important to preserve the business rules than to preserve the physical format

- System should have user-friendly front-end

- Electronic transfer
  - PKI - must be signed by role, not by name.

- ERA should have certified (COTS) systems and trainers

- System compatibility is key

- Examine the trade off between creating a system that can store a record and maintain its functionality now, compared to just storing the record and investing in restoring its functionality when, and if, it is needed

- Issue of authenticity when normalizing (i.e., physical bit stream conversion) of the electronic record

3.2   RESEARCHER SUMMARY

The main inputs from the facilitated discussions with researchers related to the user services, the system interface and the searchability requirements. Many attendees believed that NARA needed to take the time to clearly identify and define all the potential system users. This process would assist NARA in defining the user services that ERA would require. The attendees had several specific requirements for both the system interface and the searchability capabilities of ERA.

Several other potential system requirements were identified by the attendees and are listed below in a summary of the notes from the four researcher sessions. The complete notes from all sessions can be reviewed in **Appendix B, Dialogue Session Notes.**

- Good definition of the users is required to understand and define system requirements
    - Include scientists, engineers, students as your consumers

- User Services
    - Publicize what is coming online via a news letter – avoid junk emails
    - Offer both tutorial and orientation online with incentives for users to take advantage of them
    - Use Fee structure for different service requests above basic services
    - A uniform/consistent vocabulary for terms and definitions
    - A thesaurus that provides for translating from the old agency names etc
    - Keep downloads of the necessary browser plug-ins to a minimum

- System interface requirements
    - Customized interface for state government
    - Customized interface for non federal government
    - Combined interface for electronic and paper records to search on-line for both
    - Prefer uniform user interface/search features as opposed to a collection of the original diverse finding aids.
    - Include several user-friendly options for displaying records - ability to scroll and to page as well as to go to a certain page
    - Developers should consider PDF/A for display
    - Developers should learn from software packages such as Melvin and SFX.

- Search capability requirements
  - Users would like to perform a multi-step search by looking through metadata for a set of records and maybe by searching again through the set they just found. Only after the set of records is sufficiently small, a search through the actual data within records would be required.
  - Search based on content or metadata
  - Query rules for front end search
  - Receive record back within an hour
  - Ability to download a file if you want to manipulate it directly
  - Repeated searches – save ID numbers, master locator register e.g., favorites
  - Save searches in system – provide capability to let users save cookies or bookmarks in their system
  - NARA needs to learn from libraries regarding tools to search through catalogues.
  - Subject matter experts that can be contacted for help with searches should be available
  - Provide a time and cost estimate for a received search request
  - FOIA Requests – make people go through the process. Engine that filters sensitive info - (location of power plants) then point to authority figure or process that the user would have to complete
  - Research requests with rules
  - Ability to fire an agent (a spider) that will slowly crawl through the data inside the records while narrowing things down with metadata (categories) searches
  - Save searches on the system so that searches could be shared (if the creator is willing) with other researchers
  - Prepare for "denial of service" attacks that are difficult to handle: e.g., a large set of individuals directed by an interest group requesting large sets of documents at the same time
  - Users want a sophisticated search engine that provides for wild card, truncation, and fuzzy (i.e., the system should be able to suggest key words close to the ones entered by the users) search capabilities
  - SFXIT.com by Exlibris has good model for search
  - Ability to save searches on local system via some browser mechanism without any information saved on the ERA server and without the necessity of user registration
  - Ability to drill-up to find a relevant record through listings of the records in the same group/series.
  - Consider issues with future search tools if you did not have (tag) enough metadata fields when you stored the originals?

- Selective redaction of records
  - For example records that contain things like SSN, mother's maiden name, Military ID, DOB, etc.
  - Need to keep flexibility as to which records will be considered private and not releasable to the third parties at all

- User Registration Requirements
  - Proof of identity
  - Electronic signature
  - User name, password not connected to person
  - Identity management and access controls are very important
  - More stringent registration process for access to classified documents

- Develop Public/Private partnerships (Friends of the National Archives)

- Internet Standards
  - Do not optimize ERA for broadband
  - The top 4 or 5 most popular Internet Browsers should be supported

- Keep separate unclassified and classified systems

- Collect and store as many statistics as you can
  - Can help with funding
  - Inform NARA which paper records should be digitized first
  - Help tuning the technical aspects of the system

- Database Management
  - Do not flatten relational databases
  - Create relationships between the databases you preserve

- Solicit user input when conducting ERA's Design Reviews

- Use cataloging schema that uses standards like MARC

- Train-the-Trainer approach
  - Utilize what is already in place e.g., training folks at the public library and corporations.
  - Trickle down mechanism for training

3.3   GENERAL SUMMARY

The Dialogue Session discussions also generated several important concerns and suggestions from the attendees not specific to either one of the defined user groups. The following list is a summary of the general comments extracted from the notes of all facilitated sessions. The complete notes from all sessions can be reviewed in **Appendix B, Dialogue Session Notes.**

- ERA will have to evolve with technology and user requests

- Need to address the conflicting needs of Homeland Security and Privacy Protection

- The states are desperate for a cheap system that also helps them - they cannot use ERA if the overall cost is too high to them. Their political support is important.

- Utility of ERMS systems allows any user to become records manager - procedures must address this problem.

- People only manage their electronic records when they are forced to

- The IT officers must work with record managers and record officers – the only way to develop the 'best' system

- A lack of understanding of what records are clearly exists across all agencies and individuals

- There is a sense that with network backups, "all my records problems are gone." Distinction between "backup" and "archiving" of electronic records is not well understood.

- Seems to be a lack of guidance and budget for electronic record management and archiving

- Electronic Record Center
  - Critical as a requirement. Cannot afford not to have it.
  - Major issues and cost in creating a local disjoint approach
  - Needed to propagate a solution through a central source. NARA needs to play a leadership role in promoting standards and "best practices"

- Cannot afford not to do Temporary records. They provide a cost justification so that CIO can be sold on the need to invest into ERMS

- Clearly explain the agency role and the NARA role with ERA
  - The system appears to be looking at the record lifecycle, but part of the issue is coordinating workflow. So what are the agencies' responsibilities and what will ERA do.

- Interact with international archives and international organizations located within the US, such as the UN

- Some ERA skeptics do not believe that it will be possible to develop a system to archive and retain functionality over the long term

- Appears to be a host of different and disparate records management projects that are all ongoing

- Major problem is the cultural change that such a policy creates

- Records and Archives Management Systems – all agencies use them but they are all different and separate systems

- NARA should certify agency training

- Need RMS in agencies before ERA is set up

- ERA has to be simple to implement it across agencies. Simplicity and tools support are keys

- ERA also requires a common terminology among record managers to be understood and used properly – use a targeted assisted program to develop consistency across agencies

- An OMB circular is needed to mandate saving electronic records.

Appendix A: Summary of Dialogue Session Attendee Categories

Attendee Description Category	Total per Category	% per Total Responses
Federal agency personnel, civilian or military	9	60.00%
College or university staff or student, business employee, researcher	0	0.00%
Educator or student for K12 grades	0	0.00%
Genealogist, family historian or genealogical researcher	1	6.67%
Professional or non-profit educational organization	0	0.00%
State or local government agency personnel	4	26.67%
Foundation or Friends group associated with NARA	0	0.00%
General public	0	0.00%
Other Categories:		
Fed agency contractor	1	6.67%
Not Specified	0	0.00%
Editor	0	0.00%
Private Archivist	0	0.00%
Total Responses	**15**	**100.00%**

Table A-1: Atlanta: Summary of Attendee Categories

Attendee Description Category	Total per Category	% per Total Responses
Federal agency personnel, civilian or military	9	30.00%
College or university staff or student, business employee, researcher	1	3.33%
Educator or student for K12 grades	0	0.00%
Genealogist, family historian or genealogical researcher	1	3.33%
Professional or non-profit educational organization	4	13.33%
State or local government agency personnel	8	26.67%
Foundation or Friends group associated with NARA	0	0.00%
General public	0	0.00%
Other Categories:		
International Civil Servant	2	6.67%
Not Specified	3	10.00%
Editor	1	3.33%
Private Archivist	1	3.33%
Total Responses	**30**	**100.00%**

Table A-2: New York: Summary of Attendee Categories

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix A*

Final

Attendee Description Category	Total per Category	% per Total Responses
Federal agency personnel, civilian or military	3	37.50%
College or university staff or student, business employee, researcher	1	12.50%
Educator or student for K12 grades	0	0.00%
Genealogist, family historian or genealogical researcher	1	12.50%
Professional or non-profit educational organization	1	12.50%
State or local government agency personnel	0	0.00%
Foundation or Friends group associated with NARA	0	0.00%
General public	0	0.00%
Other Categories:		
Federal Agency Contractor	1	12.50%
Not Specified	1	12.50%
Editor	0	0.00%
Private Archivist	0	0.00%
Total Responses	**8**	**100.00%**

Table A-3: Los Angeles: Summary of Attendee Categories

Attendee Description Category	Total per Category	% per Total Responses
Federal agency personnel, civilian or military	19	70.37%
College or university staff or student, business employee, researcher	0	0.00%
Educator or student for K12 grades	0	0.00%
Genealogist, family historian or genealogical researcher	0	0.00%
Professional or non-profit educational organization	2	7.41%
State or local government agency personnel	3	11.11%
Foundation or Friends group associated with NARA	0	0.00%
General public	0	0.00%
Other Categories:		
Information Manager	1	3.70%
Not Specified	2	7.41%
Total Responses	**27**	**100.00%**

Table A-4: Chicago: Summary of Attendee Categories

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

Appendix B: Dialogue Session Notes

B.1   Atlanta Producer Notes 1

- Can't even read it. Migration as an approach is OK. No value (in term of authenticity) in preserving the old forma. Data can be re-manipulated by business rules must be preserved
- Example of old DB of voting record and pattern
- PDF engineering drawing. TIFF, scanned. Come from contractors, not gvt. No electronic transfer. Process is still paper based (stamped date, signature on drawing). As long as we can ingest engineering drawing and bring it back for access, it's OK. Can send binary file to printer for print out.
- Email - PDF as email. Documentum with VB front end. Save mail records twice a day. Good feature of Documentum is querying metadata, not workflow
- Metadata - Organized by keywords, file #, subject. Biggest problem is not to define MD but to get users to enter them. Incentive is key. User needs to see the benefit of having these items organized and having the info at their fingertips.
- Search based on content - People will expect it. Getting record back within an hour is OK. Download file if you want to manipulate it directly yourself.
- Record management - People don't' care. Except for lawyers, the court, auditors. Key is to get IT management to embrace record management.
- Mechanism to delete record (make them non-permanent) once they are deemed permanent is important. (What was the rationale for this?)
- Electronic transfer. PKI. Must be signed by role, not by name.
- Notion of sensitive document. Not just secure. Example is image of signature of important people.
- Need to scale down
- Fewer record type. Need to have customized interface for state government. Different priority. Different type of access. Simple and cheap. Scale down as well as scale up. We always think in term of "tidal wave" I realize that in the field the states are desperate for cheap system that helps them. They can't use it if the cost is too high. Their political support is important. Most technology works from bottom up, not top down.
- (User) Denial of service attack to inundate the system with frivolous requests. Throttling system performance. Governing resources. Ex of professor getting students to file requests.
- Ability to redact record on the fly
- Kentucky. Birth data record on CD. Available for sales
- Carter library example: very old system. Search on Andy Young and get no hit. Neither is Andrew Young, or Young, Andy. Finally try "Honorable Andy Young" to get the 57 hits. Issue with "soft" error or flexible search. Flat file as DB type!
- No concept of "schedule" (i.e. removal of records)15-20 M to 35 M imaging record (environmental)

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*
Final

B.2   Atlanta Producer Notes 2

- Scale Down - After categorization, where to put records?
- Email types
  - Lotus Notes—dbase environment (electronic organization)
  - Does not "look like" email
  - Eudora popmail-→ Outlook
  - Web browsers--→ Outlook
    - opposition to standardization
- Databases
  - RDBMS common
  - other types?
- Native database formats necessary?
  - prefer migration WITH functionality
  - dbase snapshot as relates to records
  - Administrative use of database
    - preserve use rules
  - Reports may crudely capture a dbase over time
  - Spaceflight docs in PDF
- [Lotus has output feature of email to PDF]
- [use of Lotus & Documentation for data association]
- Imaging types
  - TIFF grp 4foro engineering drawings, not yet using CADD at Marshall SFC
    - CADD to scan
    - workflow stamps and signature (paper not electronic—new stds for CADD drawings?)
    - automatic systems may use PKI signatures periodically
  - Metadata
  - RMAs: metadata ans query tasks?
- _____ docs
  - incentive for users
  - friendly front-end
  - provision of search tool for systems
- Searching for Records
  - metadata or content search
  - query rules for front-end search
    - expectations
  - downloadable files→component level
- Sense of understanding of ERM problems?
  - many don't want to know
  - lack of understanding of what records are
    - sense that with networked backups, "all my records problems are gone."
  - individual records creators use various RMAs
  - procedures not in place

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

- legal and audit need to be partnered with mgrs/creators (also security)
- authority of archivists
- mgmt understanding necessary
- lack of guidance/budget for ERM
- Which record classes are most important to implement first?
  - PDFs
  - mission driven
  - email, scanned images, CADD/Engineering
    - std and interoperability of CADD
  - Audio/Video digital files
  - GIS
  - digital image format - photography
    - capturing
    - permanent storage
    - format
    - clarification for transfer to NARA
  - How can system help us set priorities?

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

B.3   Atlanta Researcher Notes 1

- Make sure that you have a specialized User Interface (homepage) for your users at State Agencies and even large Municipal agencies.
- We would like to receive an e-mail with the summary of the meeting.
- Make sure that your AOL, Netscape, dial-up and other low-end users are not forgotten. Do not optimize your system for broadband.
- Yes, we understand if you let us search through the metadata first before letting us search through sets of Records inside of them, but it would be nice to be able to fire an agent (a spider) that will slowly crawl through the data inside the records while we are narrowing things down with metadata (categories) searches.
- We would also like to see a Master location register so that once we find something, we can have an ID for that records and can pull it up without going through any more searching.
- You might consider letting consumers save their searches on their local system via some browser mechanism without any information being saved on the ERA server and without the necessity of user registration.
- Do separate the unclassified and classified systems. We are afraid of unauthorized users.
- If you decide that the document is sensitive, but can be released under certain circumstances, ask the user to contact an archivist or to register using more sophisticated credentials.
- Let us know which popular non-digital documents are being digitized or at least catalogued within ERA as it happens.
- Beware and prepare for "denial of service" attacks that are difficult to handle: e.g. a large set of individuals directed by an interest group requesting large sets of documents at the same time. Also remember that the prison population is famous for frivolous requests. Slow them down because if you build it, they will come ☺
- You can charge for services beyond basic if you are reasonable about it. For example you can charge for extensive mediated searches or for expedited services or for shipping CDs.

- Remember that you might have to selectively redact records that contain things like SSN, mother's maiden name, Military ID, DOB, etc. New legislation might be passed and you need to keep flexibility as to which records will be considered private and not releasable to the third parties at all.
- Make on-line orientation to the Archives mandatory or at least provide incentives (like extended service discounts) for people who take it.
- Genealogists would love to see a detailed on-line tutorial on the use of the archives.

B.4   Atlanta Researcher Notes 2

- Partner with libraries (searches) – lessons learned & best practices
- User Registration
  - Proof of identity
  - Security
  - Electronic signature
  - Privacy
  - Classified records

- Popular records online
  - Vote on records online
  - Snapshots
  - Finding Aids, list of everything
  - Online ordering, email, mail

- Search Capability

- User Interface State Gov't

- Non-Federal Gov't Interface

- Major Internet Providers
  - What are they doing
  - Dial up internet

- Search Both: Metadata and actual Records

- Master Location Register - Location ID

- Bookmark the search (on user's system) – Instead of ERA system saving it

- Level of Service
  - FOIA Requests
  - Research Requests with Rules
  - Basic contract services
  - Fee for services above basic services
  - Orientation Online
  - Tutorials online
  - Benefit of going through orientation
  - Mandatory orientation?
  - User name, password not connected to person
  - Public/Private partnerships (Friends of the National Archives)

B.5   New York Producer Notes 1

- NY State Court system - Develop certified standard for record, schedule conversion to microfilm then to e-record! Librarian to create index.
- Contractor does the CAD work. Want to transfer to government in electronic format but the government does not have a way to take it.
- Interface with COTS tool for presentation. Important to support variety of tools. However, if the only thing you can do is this, they will take it.
- Email. Need filing. Filtering - built in scheduling. TRIM. ERMS product. Preservation Vs operation.
- Attachment must be preserved. Formula as well as format.
- Electronic Record center - Critical as a requirement. Cannot afford not to have it. Huge cost otherwise to create local disjoint approach. Needed to propagate solution. Scale down. ERA at agencies. Organization has tough time keeping track of regular operation, let along record management,
- Email as early test case. - CIO/COO understands the need to deal with them. Will most easily get support.
- Access charge requirement - to calculate charge, system has to have ability to associate pricing value with record volume. Charge commercial customer.
- Usage statistics for record as a performance data. What about warehouse of performance and record access statistics? Built in tuning.

B.6   New York Producer Notes 2

- Although dealing with CAT drawings, GIS and email we have no system in place for archiving them
- Interested in knowing if there are any plans on dealing with UN agencies, where the host nation is US
- Are different international archives talking with each other?
- The documents we need for the long-term (hundreds of years) are mostly paper
- ERA skeptic – do not believe that it will be possible to develop a system to archive and retain functionality over the long term
- See advantages of electronic filing over paper filing – recovery from misfiling paper records is impossible
- If we go to electronic, will we lose paper records?
- Will paper records be scanned to be archived electronically?
- Deal with investigative files that are all paper –nothing electronic
- See a limited creation of electronic records
- At the point of dealing with electronic records archiving and want to know what to do
- Looking for guideline on electronic archiving due to OPRA – now easy for people to request access to all records, therefore, looking to do so electronically
- Little communication between IT and decentralized records departments, therefore, generally poor record keeping
- Job is to help guide courts to archive records – set up one of the first electronic archiving systems, where we have scanned many permanent files (will etc) from microfilm and now keep them electronically. Many courts doing their own thing and thus have to fall standards (provided Fynnette with a step by step guide)
- Not gone to electronic archiving and not planning to
- Some standards have been set (DOD) but major concern is that agencies will come up with their own systems/standards – how will NARA deal with all of them?
- Appears to be a host of different and disparate projects that are all ongoing
- Selected a record type (email) as a starting point.
  - Email has lots of different standards, involves attachments and has interesting metadata
  - Issues with different generation email systems
  - Is email actually a record, as per the definition in the presentation? (General consensus of yes)
  - The content of the email should dictate if it's a record to be kept or if it can be deleted
  - Impact of time – some systems have automatic deletion after a period of time – the way to keep it as a permanent record is to print it out
  - How can the process of archiving and disposing be managed?
  - Is there a system that actually archives emails?
  - Issues within different industries – SEC require financial institutions to keep all emails for 7 years – therefore clearly a record
  - Printing it out is the only option unless a true ERMS is in place

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

- At the office level, most have no idea of how to manage and archive electronic records using file directories, hierarchies etc
  - Such a system is potentially easy if the exchange server is set up properly
- Agencies/firms tend to manage computer systems vs. managing information (MIS vs. Record management)
  - Keeping the customer (user) computer system working takes up all the time
- If no real archiving system in place – the only option is printing, but if the paper records filing are not kept up to date the records are being lost on people's desks….
  - Computers have replaced secretaries and office assistants – therefore the available labor for filing paper records is diminished, but the expertise / systems for electronic filing do not exist
- Generally people do not manage their email – some of the junk should just be deleted.
  - Filtering the email should be a requirement – create intelligence in the system to help
- Printing by law is not a solution for NARA – lose too much information
  - Clear issues of volume that needs tool support
  - Some email does need to preserved forever
- Need training to be provided for what emails need to be preserved and which can be deleted – what is a record in terms of email?
  - People will store all their records as they have the space (memory) to do so – even though many of the records are not actually needed
- Could incorporate required archiving information into the system – make users complete information at the time of creation / amendment – details can then be used for tracking
  - Can you impose this on users?
  - Could use required fields
- Main issues are training and technology – but also authority.
  - Many of the solutions will not work unless users are forced to use them.
  - How does NARA get the federal agency to comply?
  - POLICY ISSUE – in the parking lot
- MS Office has a tool to categorize records as they are created but nobody uses it –
  - Could make categorizing tool pop before saving the document.
- Major problem is the cultural change that such a policy creates
- Records and Archives Management Systems – all agencies use them but they are all different and separate systems
- For User Training – need to justify to them the benefits – why is it in THEIR self interest to actually use the system to get their records from their email system to the archive system
- Records Management Systems work with email systems such as wang, Lotus Notes
- Interesting idea to impose record description pre electronic record creation, but dream world.
  - Most projects already behind schedule and over budget and this is just another layer.
  - Most do not have an integrated RMS to start with
- What about the complications with attachments?
  - Possible solution is collaborative work using version control – centralizes documents for ready access

- Not acceptable to mandate this type of system
- What is relevant within the attachment – the image, the formulas, ability to change it?
- What are the final versions of the attachments?
- Who keeps the attachment – the sender or the receiver?
- Need RMS in agencies before ERA is set up.
    - Is ERA for archiving permanent records only? If not all encompassing of the lifecycle, a RMS is needed to schedule and appraise records before they are accepted into ERA.
    - Is the scheduling and appraisal remaining as an Agency responsibility?
    - What format should be sent to NARA?
    - Are Records Managers empowered by NARA or do they have the ability to take control of the process?
    - Do they have the budget to do so?
    - There is recognition that the problem exists but there is no obvious solution beyond checking every user's computer to see if they are correctly archiving and disposing of their electronic records
- The current work being done by contractors is forcing some agencies to use more and more electronic records
- Within DOD possible to enforce use of standards, but in my agency there is 6000 employees with low level of computer literacy and we have no power to force them to start using technology
- The idea is to correct/improve the workflow for electronic record archiving and not do the jobs of the records managers for the agencies (some found the introduction presentation misleading on this point)
- What about system compatibility – this seems to be beyond a "major issue" as no option to dictate to the agencies
- What about accessibility?
    - A record is not useful unless we can access it without:
        - Losing functionality of the record
        - Losing the metadata and content
- What is the trade of between creating a system that can store a record and maintain its functionality now, compared just storing the record and investing in restoring its functionality when, and if it is needed?
- What about databases?
    - AAD – flattens everything down to tables.
        - Issue of authenticity when normalizing the data.
    - AAD if only a collection of tables what happens to the relationships within the data – are they not lost? What about the capability to print reports?
    - What is the value in storing the raw data? Does not seem much point in storing it if the usability is not preserved.
- Issues of retaining:
    - Primary value – content
    - Secondary Value – evidential
    - Tertiary value – information

- What is permanent? Records should be reviewed every 5 years to reconfirm permanency
- Most electronic information is instant but some holds long term value and that has to remain the main focus
- Hardware and software issues – take it out of proprietary nature to generic e.g. XML
- If I gave you an 8-inch disc with such and such a file on it, how would you access it today?
- Will ERA just deal with permanent records? Or temporary records also?
- This seems to be an unmanageable task!
- What is permanent needs to be predetermined
- Other issues of security classification and privacy
- Cost justification – should be cheaper for NARA to go out and develop an all encompassing system than for each agency developing its own system
  - A centralized system requires an agency unified approach
- With DOD standards, software vendors have jumped onto the standards – NARA could try to exploit this concept in setting standards
- What is actually being done with current information – the Clinton emails, [Being used as a testing ground but actually printed out]
- NARA's biggest issue is the resources available and its will to do it.
- What will the NARA budget be for the project?
- Agency funding – NARA could look at certifications for agencies.
  - NARA could certify agency training / systems
  - Contract for audit to maintain standards
  - Agency has to build to the standards that are set and monitored by NARA
  - Need to make the best use of the resources to meet requirements
- What is the implementation approach – looking for an early success to gain credibility?
- Need both the code for the operating system and the bitstream data for excel – the excel data is meaningless without access to all the applications tools

B.7   New York Researcher's Notes

- We would like to see a combined interface for electronic and paper records so that we can search on-line for either.
- The top 4 or 5 most popular Internet Browsers should be supported.
- Staged search (first through metadata and then through the set of record of interest found in the first step) is fine.
- We want abstracts and descriptions on a level more granular than a series (depending on the series).
- We like the features of the FileMaker Pro search engine for example.
- You should save some searches on the system so that these searches could be shared (if the creator is willing) with other researchers.
- We prefer a uniform user interface / search features as opposed to a collection of the original diverse finding aids.
- A uniform / consistent vocabulary for terms and definitions is good, but we would also like to see a thesaurus that provides for translating from the old agency names etc.
- Do not abandon your current efforts like ARC. Expand on these efforts.
- You may charge for certified record copies.
- You may charge for large electronic downloads. Maybe per megabyte. Similar to e-book: a free chapter, but not the book.
- You may charge for expedited service, i.e. moving people up the service queue.
- You may charge for special extended services but you have to be careful about that: the process of determining the fees might be too complex and costly and thus not worth it.
- You might try to differentiate between a private citizen and commercial entity customers as far as charging for services. The commercial entities will try to protest, but it does not hurt to try. For example FOIA does not include commercial entities as customers so you might get away with charging them for records especially in large volumes.
- We would like to see different User Interfaces provided for different groups of users. This customization can also depend on the level of access granted to users.
- Collect statistics on usage. Keep as many statistics as you can. This will help you with getting funds and will tell you which paper records should be digitized first (see 26 below). The statistics are also important for tuning the technical aspects of the system.
- Users want a sophisticated search engine that provides for wild card, truncation, and fuzzy (i.e. the system should be able to suggest key words close to the ones entered by the users) search capabilities.
- Make sure that you have several user-friendly options for displaying your records. We want to be able to scroll and to page as well as to go to a certain page. You should consider PDF-A for display.
- We would like to have available a list of subject matter experts that can be contacted for help with searches. We like Google's designated researches list for example.
- Add value for the users who register by providing them with updates of new records and/or series. You should also have a monthly Newsletter which you might be able to charge for.

- Chats might be unnecessary due to a large number of general interest listservs, but the students might appreciate the medium. Also a lot of revenue might be generates through chats if you have to keep an expert available 24-7.
- Possible downloads of the necessary browser plug-ins are unavoidable, but try to make them as infrequent as possible.
- Have a help line service (ask an archivist), but not 24 hours a day.
- Provide a time and cost estimate for a received search request. That would greatly reduce the frustration of expecting the services too fast.
- In general, do not forget that some users might have slow modems.
- If somebody wants a digital copy of a paper document, you might be able to charge for digitization and thus increase the number of digitized documents that can later be provided free of charge.
- We would like to receive an e-mail with the summary of this meeting

B.8   Los Angeles Producer Notes 1

- Project records, about 40,000 sq. feet. Not much electronic yet. Voyager records. Working on initiatives to convert to electronic records. Use "DocuShare" tool from Xerox. Lack of NARA guidance. Currently part electronic, part paper.
- Genealogical data. 1930 Census. Web access - Discussion on privacy issue. Access control at web/login. Registry of user. ID/password. Based on individual.
- Search mechanism - with some sort of authority based priority so as not to get a lot of false alarm. <Good requirement>. Key word search is not authoritative enough. OCLC WordCat. SFXIT.com by Exlibris has good model for search. Breaks data sources down in term of data types. Organize what you have access to. Categorize into what are book sources, what is on line, etc. CalTech implement it.
- Bankruptcy Court for LA District - Moving to electronic filing. PDF with email confirm. No authentication. Anybody can file bankruptcy for anybody else!
- Need to support BOTH a paper-based process as well as an electronic process.
- Issue of searching of multiple data sources - Finding Aids on how to aggregate these searches so that one can handle large number of data sources all at once. Need to be able to do "search within a search". Subset requirement.
- How to handle data that are sensitive but not secret. Divorce data.
- Save/Autoload your record. Save your search.
- EnDeavour
- Privacy issue
- High level of efficiency.
- All user audience.

B.9   Los Angeles Producer Notes 2

- Records archived by Project (i.e., Voyager, Hubble) and consist of drawings, images, and administrative documentation. Records date back to the 1940s. JPL has an archives and its users are JPL scientists and the public. Currently, the archives are not electronic. Issue: How to capture ERs and how to transfer them to NARA. JPL uses a tool by Xerox named DocShare. As projects close, JPL captures records into an inactive archives then once records are appraised, records will be moved to an active archives then finally to permanent archives.

- NARA Regional Records Director, predominance audience is in late 20s on up to 80 or 90. Most of the attention is on textual records – have contractors, native Americans, and bankruptcy records. Challenge: parts of files are electronic, some are textual, classes are mixed because courts haven't insisted that filing be done digitally. See site at archives.gov/office of regional services/Laguna Niquel/. Holdings listed in NAIL, now in ARC.

- Internet & web, paid subscription services, 1930s, ancestry.com, genealogy.com. Process is interactive – start most recent and trace backwards. Hoping to find out what will be available in ERA.

- Name, Index, Misspellings, Nicknames must all be tied together

- Administers email. Interested in proper access methods to manage people – identity management and access control. Exploring method to use a single directory to authenticate access. Access control – restricted information that certain people inside JPL can access but access denied to general people.

- How done now – historically islands of access so system has rights. Want to move towards one centralized place to grant access. One registry of users with registry information. Access to groups and avoid by person because highest maintenance. Sometimes, license can equate to single person.

- Users: Faculty – access to gov't information, Students – MBA programs & business students, Administrative users – census information. Like methods used by the Library of Congress – large schema to find people, places, etc. Searches – keyword searches can fail so that it is important to look at access and retrieval issues. Catalogued – authority searching is good, use schemas. For example, LC creates an authorized record for a person including birth & death dates. Need records to indicate who individual was, places, subject heading. Need good subject schema: Name, subject, place. Gov't needs cataloging schema that uses standards like MARC. Use the NARA as a starting point for all the agencies – ERA should track to schemas used by agencies and LC. MARK records URLs.

- Tapes in storage but can't be read so losing data. Military personnel data and corporate archives for RAND Publications. Library is responsible for preserving records forever, problem with mold & fungus.

- Case Files going electronic filing to courts. Focus is compliance with EAO. Cast Management System Database – temporary server and permanent server. Building a database to handle electronic filing. Interested in presentation media and formats

- Use PDF, registrar, validated, receive email, confirmation and base

- The problem is that courts require original signature and old filers send paper to the courts and the courts scan it but NARA needs PDFs. For courts, mix methods of filing so get mixed media.
- PKI protects people from identity theft but it doesn't really.
- Authentication is a huge issue.
- Identity protection and privacy is also a huge issue.
- We are going to try to go to electronic filing.
- Disjoint between requirements of the court and the requirements of a creditor. If use entire case file, must use NARA and state level (paper).

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

B.10   Los Angeles Researcher Notes 1

- Identity management and access controls are very important. Identity theft is a problem.
- Records must be hierarchically structured and searchable by metadata. Keyword search is not adequate. You need to use some standard like MARC, but maybe a more manageable one.
- Staged search (first through metadata and then through the set of record of interest found in the first step) is fine. In fact, it is very much desired. We do not want to just add more criteria and rerun the search. We want to be able to search within the result set of a previous search.
- We prefer a uniform user interface / search features as opposed to a collection of the original diverse finding aids. Please learn from such software packages as Melvin and SFX.
- It would be nice if your search result came ordered from the most relevant records to the least relevant ones. We would also like to see a "more hits like this" feature.
- We want to be able to drill-up, i.e. if we find a relevant record we want to be able to see some listings of the records in the same group / series.
- Maybe by 2007 you can assume that everyone will have broadband. Sometimes it is OK to stop supporting old technologies. Maybe each Post Office will have a kiosk.
- Your users are getting more educated and savvy. Make sure that your system can evolve.
- How are you going to address the conflicting needs of Homeland Security and Privacy Protection?
- Do save user preferences and save searches and search results.
- You should be able to charge for any service that uses up a lot of computer resources.
- Are you consulting information brokers? Are you learning from them how they charge for services?
- Do not flatten relational databases. In fact we would like to see more relationships established between the databases you preserve.
- Provide on-line help. Automate as much of mediated assistance as possible – people are expensive.
- Use your natural partners (librarians, federal and state records managers and archivists etc.) by training them to provide assistance to their customers and the public.
- Solicit user input when conducting ERA's Design Reviews.
- Do not forget to include scientists and engineers as your consumers.
- Consult media technologists ("news people") on how they manage their archives.
- We would like to see different User Interfaces provided for different groups of users.
- We would like to receive an e-mail with the summary of this meeting as well the meetings in other cities.
- How are you going to be able to use future search tools if you did not have (tag) enough metadata fields when you stored the originals?
- How do you preserve records permanently?
- How do you make sure that e-mail threads are preserved?

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

B.11 Los Angeles Researcher Notes 2

- Searches: Two tiered approach: First, metadata (indexing info, descriptions) then once limited, use second, search records themselves…is this a viable approach?
- Look at existing models of information databases, like Boolean mechanism, need good description of ways users can search, like search within a search – think this is the next generation of searching. The challenge is how much information search is enough. This depends on the user – undergrad vs. Doctoral search. Comes down to defining users really well.
- Would you like to save your search – save your "save set"?
- Need a mechanism to find out where information is stored. If have to go there to search – less preferable.
- SFX – ILF system. URL Standards – allows users to organize their search. This is a new concept in libraries. More info at SFXIT.com. The user doesn't have to know all databases (dbs) – on search, results bring back dbs – goes across all kinds of dbs and organizes results per user specification.
- NARA is a member of the DLF. Access control filters out search.
- Store – Will not change what you capture. Search will change. The challenge is how to store info. So it can be searched and found. No longer case to focus on content & searching on content. Searches based on content – i.e. scanned image. Technology now available for searching on images is rudimentary. There are no standards – like web base Internet. Evolve it overtime – in the beginning, won't be able to capture all content.
- Email – How to search email?
- How are you going to store data permanently?
- "Preservation" – still in research. Four approaches: 1. Corporate approach – slow migration, pro-easy to do, con – lose info? Authenticity question. 2. Microfilm – scanning – treating like paper base 3. Emulation – running program to view record- many technical issues. 4. Reduce to a standard format. Depending on record type, will require one or more approaches. Requirement: must save original file.
- Increasing challenge: Preserving databases and animation.
- Portal or locator to identify searches, then narrow it down to do record by record search? For email – several standards.
- Issue – hard format because so flexible, tedious to capture metadata. Question: how to capture massage thread?
- Search – how to capture thread to bring back all email messages?
- Standards for metadata tags? So can search up?
- ARC can drill up.
- Browsers – can not assume everyone won't have broadband.
- Yes, but only for a couple of years. By 2007, can assume people have broadband access.
- At some point, IT has to state what is too old & can't be supported anymore – then use kiosk for central access (at libraries, for example).
- Capability to search on the side? Download set of search to search later? Inside of Gov't – sensitivity, some records are sensitive – privacy issue. How do you deal with this issue?

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*
Final

- Legal issue – ask citizen to waive privacy. Will be increasingly important to protect privacy in protected at least while they're alive.
- Once NAR gets recors, privacy not an issue, need to work with crating agencies to make smooth transition.
- JGL – who can see what? Define sets of attributes, define who cans see them.
- Problem – can do with small number of data and users but if everyone has their own privacy level, this is more difficult.
- Need to have a flexible engine where we can put these key words and then tag the information.
- May have to reformat records to put new tags or indexes on them as ERA grows.
- MARC is very flexible this way. Classification is endless.
- May provide several interfaces (portal approach) and have capability for users to customize it a bit. Sound reasonable? – Yes. Saving searches? – Yes.
- Identity – how will ERA handle?
- Depends on the customer – more info given, more service a customer can use.
- What's legitimate research? How the gov't looks at this question can be a freedom issue going on.
- If have users taking up a lot of time – NARA trust fund to charge for support and expedited services.
- DBM – How envisioned? Grant access to dbs or users have to download entire dbs?
- Issue - 1. History of search in db versus final piece of info. Desired. Database – Presearch flat file capability or cloning all dbs (not realistic).
- Merging dbs?
- At individual level, only want access. Corporation will want merging, data mining. Our mission is to presearch the db, not the aggregation of info. Approach.
- Support – want online help as a minimum.
- Take Train-the-Trainer approach & training folks at the public library. Think about what's already in place and utilize this. Look at corporations, train them. Repository libraries – state level. Trickle down mechanism for training.
- Mediated search – contact with archivist?
- Google – more hits like this – nice to get in ERA. People time expensive – need to automate people time.
- General life cycle? Once you get your design – come out again and get specific feedback.
- April – Proposal in April – Design period – Design proposal not out for another year. Want to set up mailing list. Hope this is the beginning of the Dialogue.
- How work with agencies now?
- RMI – policies for agencies now. ERM – initiatives with E-Gov and new policies.
- AAD?
- First step.
- AAD doing things now in the meantime until ERA.
- Send us comments about AAD.
- Two groups to add to the user groups: scientists and legal.
- How did you do your media coordination? This is a small group here.

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

- How can we improve our outreach?
- Proximity – make it easy for users to participate

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

B.12   Chicago Producer Notes 1

- DCMA Defense Contracting Management Agency. Contract record. Folder of files. Moving to electronic.
- US EPA HW/SW issues. Lotus Notes. 30-day retention rule. Backup, not archiving.
- EPA Central Regional labs.
- Admin Office of Naval Research. Contract file moving to electronic. PDF. No PKI. No tracking of changes.
- GSA.
- US District Court
- EPA Central Regional labs. Data in paper. Moving to electronic. Data source unknown and do not need to preserve sources. Only result.
- Admin Office of Naval Research. Contract file moving to electronic. PDF. No PKI. No tracking of changes.
- GSA. Lotus Notes. 60-day retention program. Use either Lotus Note provided archiving capability or Plug in program to save email. No filtering/no classification.
- US District Court: Email. Litigation. Record schedule. Benefits administration.
- Corps of Engineering. KISS. Lots of different engineering format. All official email have a // tag and get printed to paper for saving. Financial management system is Oracle DB. How to archive.
- Email privacy issue. We cannot read every email. Same as EPA Trade Secret.
- SSA. Outlook Exchange. Save .pst files. Scan paper document.
- How to separate the context of the email topic from the media? I.e. if a file has office docs, email and images, they should be grouped under a common record series, not that each email file is a record series?
- Grouping of different record types under a common logical grouping. Where to draw the line?
- NARA to help. Get vendor to add hooks into system, for archiving purpose. Promote standard. Certification process.
- How to recognize an official copy of a record.
- Archiving web pages? Open research.

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

B.13   Chicago Producer Notes 2

- What is a record? And how do we know what needs to be archived? Especially in respect to email.
- How will ERA address the multiple programs being used today, but will be obsolete in the future?
- From archiving perspective, I am used to dealing with paper records
- Policy is to delete emails over 30 days old, responsibility of the user to archive essential email on their local drive or by printing out
- User education is essential as they do not know what to keep and what is junk
- Interested in how ERA will manage the relationship between the IT office and the Records Manager office
- Currently archive paper records but looking to archive electronically, but not sure what the right format is – currently looking into pdf files.
- With paper records we continually pull documents from the file that have been archived - how will ERA keep track of this type of activity?
- All our data results are printed out on paper to be archived
- We send out data results to users and want to know how to keep track of what has been sent and to whom
- Definite lack of archiving across our organization and records are definitely being lost – different options for archiving electronically and there is no consistency amongst methods used by users. Due to the increased number of documents being produced, sitting on a timebomb
- Seems like every agency employee is a records manager as they have access to delete records
- Is their any guidelines on the timeline for destroying records – as far as I know it varies per agency
- Organization moving to a totally electronic application, including electronic signature, but the record office is not sufficiently involved to ensure the electronic records are being managed
- ERA has to be simple to implement it across agencies
- How does NARA expect to be able to accept all the different formats of records that can be created – how can you track maps electronically?
- Issues with the privacy act and appraising emails?
- How to capture all the information within a database or spreadsheet for research purposes
- Issues are obsolescence and migration
- ERA also requires a common terminology among record managers to be understood and used properly – use a targeted assisted program to develop consistency across agencies
- All our records are currently paper, although looking at scanning technology and software to move toe electronic archiving – issues as public needs access to the files but the files need to be access protected for those permission to view them
- Emails are deleted every 30 days – do not know what to keep
- What is the intent of archiving?

Electronic Records Archive (ERA)   *Dialogue Sessions Summary Report*
ERA Program Management Office (ERA PMO)   *Appendix B*

Final

- Archiving for lawyers is different to archiving for researchers
- The user impacts the information that needs to be captured and archived.
- Would NARA accept an email dump?
  - No – guidelines exist as records need some form of organization to be part of a record schedule.
  - Dump is useless information as difficult to search and relate.
  - The long-term archiving requirement dictates that records need to be organized to be understandable 50 – 100 years down the road
- What about a pre-template for emails?
  - Easily used in all applications, as long as template is consistent and does not change
  - What about storage capability?
  - Legal aspect of letting the public know how to correlate the data
- The key characteristics to be preserved for future access?
  - Basic – sender, recipient, date, text
  - Policy issue – will the user always know what is being captured
  - Probably don't realize what is being captured today!
  - Process needs to be able to make sense of any attachments
  - General consensus on the minimum requirements
  - Text can be converted easily but the attachment is much tougher as involves different software
  - How do we store Internet links within emails?
  - Records needs content, context and structure – remove it from the email software and lost its context
  - Need to preserve the chain of emails – cannot segregate the email as a unique records
  - What about references in emails to other documents – we have emails referring to archived boxes of paper –
    - How can access for research be linked if the email is not archived with the documents?
- What are the expectations of staff managing own records as part of ERM?
  - Paper records – yes – print it out and put it in a contract file
  - Electronic records – not sure if it can be managed this way
    - 50 years from now researchers would want to look through a project file and not different email accounts to follow the flow of the project
- What about financial databases that relate to multiple projects
- Using pdf filing system rather than paper files also creates appraisal questions
- NARA needs to set standards for vendors to develop software to those standards for agencies to purchase standard software to be compatible with ERA - a certification process for vendors
  - NARA has endorsed DOD5015.2.
    - Will this make it easy for agencies to use NARA?
    - Should help but not complete solution
- How does NARA recognize the official copy of a record?
  - Will it archive multiple copies of the same email sent to different agencies
- What about archiving web pages?

- NARA is working on this – been in draft a long time as complicated issue
- Federal web sites are records

B.14   Chicago Researcher Notes

- The data supporting the initial authentication of a record must be preserved as part of the record to enable future litigation support.
- When we view a document many years from now, what are we going to compare its current presentation with to ensure authenticity?
- Most of the time we want you to save the result of a search as a set of pointers (IDs) to (of) the records for later use.
- Staged search (first through metadata and then through the set of records of interest found in the first step) is fine.
- Make sure that your system can evolve. We might want to do data mining on your holdings.
- Records must be hierarchically structured and searchable by metadata. We want a logical map of your holdings. Keyword search is not very effective. We need to be able to choose a relevant area first before conducting a time consuming detailed search.
- We have a problem with referencing another agency records as part of our records. Those agencies used to send us copies and there was duplication, but now we do not have guarantees that the external references do not disappear on us during the lifetime of our records.
- We prefer a uniform user interface / search features as opposed to a collection of the original diverse finding aids. However we might need these old finding aids sometimes to investigate the question of how the original users viewed the databases.
- What if we need to view an ad hoc report off of a GIS system several years after it was run? The originators would need to save it before ERA can preserve it.
- Your system must be versatile enough to be accessed from office computers and from tablet size ones. Do not optimize for T1 and provide alternatives (like shipping CDs) for image delivery.
- You should be able to charge for your services to recoup some costs, but you have to be careful of litigation. On the other hand, if you charge people they will conduct their searches smarter and not burden you system with unnecessarily large downloads etc.
- You maybe able to charge for mediated search.
- Establish a train the trainer program. You can establish a certification program where an ERA expert will be similar to a notary public and provide the public with the list of such experts.
- Establish a reasonable turnaround time matrix for your services and stick to it. Instantaneous demand is not a reasonable requirement. However immediate response on what documents are available might be expected.
- Maybe you can charge for expedited services. It is a double-edged sword though: people will not burden you with frivolous requests, but on the other hand they might sue you. Make sure that you provide free services to the people of low income with legitimate needs.
- If you save something in your system, make sure it does not disappear like bookmarked sited tend to do.

- No SPAM please! If you identify a user as one with a specific interest, do not put him / her on a mailing list unless he / she asks for it. A newsletter posted on the web is preferred.
- You need to establish a way for your users to get help even if you do not conduct actual free searches for them.
- How are you going to address the conflicting needs of Homeland Security and Privacy Protection?
- We are not saving electronic records at the agencies now. Agencies lack reliable back-ups and we are loosing documents.
- Our IT people do not provide an adequate infrastructure to save the scheduled records only. The IT people can save ALL records, but we cannot even afford to pay for that even if we wanted to do that.
- We would like ERA to provide Electronic Records Center functionality.
- An OMB circular is needed to mandate saving electronic records.
- We cannot provide information for MARC fields manually. Metadata should be collected at a coarser than record level or collected automatically

B.15   Washington DC Meeting Notes

Bi-monthly Records Information Discussion Group Meeting (BRIDG)

- Where is ERA in the development process? Are you gathering user requirements now? Has the feasibility been proven?

  For several years we have been working with our research partners on proving to ourselves that the target system is feasible. We are in the process of refining the user requirements. We have issued two RFIs with encouraging industry responses. We are about to issue a draft RFP. We will issue our RFP in August, select two winners and have them both architecture the system. After that we will select one of them to do the detailed design and development. We are planning to have several increments after the IOC. We already know how to deal with metadata, textual and office type of documents, e-mails. Preserving the behavior of database applications presents more challenges, as does making video independent of hardware platforms. We are going to be working on processing office type of records and e-mails first.

- You are casting a wide net now as you should, but how will you prioritize the requirements when the time to issue the RFP comes? What methodology are you using to do that?

  We are following the industry best practices. Our list consists of the requirements needed for accomplishment of the mission as well as for building a proper framework for the system as a whole. Based on this understanding of the priorities we will have the requirements separated into four groups: must have, need to have, desirable to have, and wish to have. If we are late with a must have requirement, we will delay the IOC. Need to have requirements might be shifted to later increments if necessary. Desirable to have requirements are dealt with after the first two groups and wish to have depend on the time and resources available.

- We are facing some electronic signature requirements as well as E-Government requirements to deal with information at various security levels. How do you plan to address electronic signature and other encryption methods as far as access to records goes?

  ERA will deal with data of several classifications. We will use the federal PKI infrastructure provided by NIST, GSA etc. We will also address the usual security concerns of authentication, integrity, and non-repudiation and we will deal with those at data, network, and application layers. The main goal of our architecture is to provide enough technological flexibility to address both the security and privacy concerns of our customers. The actual policy will be dictated by others. We will manage digital signatures and for a small portion of our records digital rights as well.

- You are promising to maintain relationship of records. How will you bridge the relationship between your electronic holdings and your paper holdings?

  We are aware of a significant overlap in business processes that go into processing electronic and paper documents. We are planning to use the automation of these processes provided by ERA in processing the paper documents in the future. However, we need to limit the scope of ERA at this point so that we can deliver the system in realistic time.

- You mentioned your research partners. Are they providing consultations for free? What is the final price tag for ERA?

  We have good support from the people who control funding – OMB and Congress.

  Partners are in a sense free, because they have their own funding. We are of course spending some of our money too, but the biggest benefit of our partnership is the fact that we are making those organizations aware of archival challenges.

  Our partners help us solve our problems. They get funding from sources like NSF.

- Technology changes every 18 months and yet you are promising to build a big funnel that will transform all the variety into something that will be platform independent. Moreover, your system will do it effectively and in an inexpensive way, preserve the records far into the future and make them accessible to many customers with different needs. Is this possible?

  You have just described a vision. Can we build such a system, turn it on and leave it running as is until the end of the Republic? Of course not! Technology will change but so will ERA! Just 5 years ago, my answer would have been no, it was not possible to build such a system. Today, however, technology and standardization are at a point where we see the lines bounding our problem converging to a near-future solution. Most of the vendors of the office type documents are putting the metadata hooks into their products and some of them supply transformation into a generic canonical format. But the main point is that ERA will be a living thing and will incorporate into its framework the most stable industry components available at the time.

- Make sure you include scientists in your customer list. We have a need to be able to see an inventory of all of the record holdings submitted by many of our offices at different times. Currently such information is not available. Would I be able to get such information from ERA?

  Several fields in the metadata we collect will address the origin of a record. Once the records are ingested you will be able to launch a query pulling the IDs of the records that came from different departments of the same agency. Moreover, you will be able to save such a query for future use. We will also provide mediated searches where you will get

initial assistance in constructing such a query. Based on your interest in this feature we might be able to include this kind of a query into our product right away.

As far as looking at such records still in the ingest pipeline, we are trying to build a system in such a way that the ingest verification happens relatively quickly.

Are you talking about seeing all of the records your agency submitted over a say five-year period?

Yes.

- Are educators and students included into your customer list?

  Yes. We also have several suggestions from the library community on how to use public libraries experience in helping such users.

  We are trying to create a virtual environment similar to the current walk-in environment at the Archives buildings.

- What methodology are you following in your system development? Is it sophisticated enough to lead you to the most important points that need to be addressed?

  We are following the Government best practices. For example we are reviewing NARA's SDLC and following some of the recommendation in it.

  The overall methodology is ours. We are incorporating methods from different ones to suit our needs.

  Some of these methods we are following are IEEE Software Development standards, OMB IT acquisition guidance, and SEI CMM standards.

- User comment: If ERA demands a lot of metadata per record and does not provide the automatic harvesting of that metadata, ERA we will not see too many records coming in.